

# Weekly Report

Aug 9th ~ Aug 13th

## Done

1. We come across too many problems while applying the model to wiki data.
  - a. Tried several edge significance method, failed
  - b. Filter edges, then apply Box-Cox transformation to ensure the data is normal distribution.
  - c. Using different distance measure, i.e., Jensen-Shannon index.
  - d. Due to the sparsity of talk vectors, we modify the distance computation method and accelerate 100X.
2. A modification on the projection in BotRadar, with multi-thread and C-python hybrid programming, accelerate 12X.
3. Learned basics of KOA2 in JavaScript, which is responsible of backend of BotRadar. Learned the concept of sync and async in JS, and the mechanisms behind them.
4. Minor revisions on BeXplorer.

## To Do

1. Find patterns on the new data. Write down, and make amendment in cover letter and paper of BeXplorer.
2. Begin revising BotRadar paper, consider which conference to submit.

## Paper

1. CHI16 *Unsupervised Clickstream Clustering for User Behavior Analysis*

一篇 CHI 上的可视化文章. 采用一种比较不太一样的自顶向下的层次聚类方法, 对行为序列聚类分析. 利用序列子序列作为特征, 进行聚类. 聚类时候, 每次用掉几个特征, 一层层往下. 越是上层的特征越是有代表性.

用到的可视化非常非常简单, 直接调用库, 利用可视化完成一系列的任务时也没有明显的体系的感觉. 但是可以说是做的比较详尽完备的, 达到了 decent 的要求. 但是若发在 kdd 或者 vis 可能达不到要求, 所以发在 chi 这一会议.